AN EVALUATION OF THE DICKEY-FULLER LINE OF UNIT ROOTS TESTS AND THE NORMALIZED INFORMATION CRITERIA IN A UNIVARIATE AUTOREGRESSIVE-MOVING AVERAGE MODEL

By

YANG-SEOB LEE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1989

ACKNOWLEDGEMENTS

I would like to thank the members of my dissertation committee, Professors G. S. Maddala, David A. Denslow, Mark Rush, and John S. Shonkwiler, for their support and encouragement. Special thanks are due to my committee chairman, Graduate Research Professor G. S. Maddala, for his care, patience, and financial support. Finally, I would like to thank my wife Young-Ran for her devotion and endurance throughout all these years.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

AN EVALUATION OF THE DICKEY-FULLER LINE OF UNIT ROOTS TESTS AND THE NORMALIZED INFORMATION CRITERIA IN A UNIVARIATE AUTOREGRESSIVE-MOVING AVERAGE MODEL

By

Yang-Seob Lee

August, 1989

Chairman: Graduate Research Professor G. S. Maddala
Major Department: Economics

In recent years, the possible nonstationary behavior of many macroeconomic time series has drawn attention from both theorists and empiricists. Formal testing methods for unit roots in autoregressive-moving average (ARMA) models have taken many different forms. Lately, Schwert investigated ARMA models with large moving average coefficients, which are thought to be common in macroeconomic time series. He reported that in these models many tests deviate considerably from the limiting distribution of the corresponding Dickey and Fuller tests, showing varying degrees of discrepancy depending on the model.

The current research extends this investigation to the four methods based on Phillips, Phillips and Perron, Said and Dickey, and Solo, and examines the normalized information criterion by Ozaki and similar extensions for unit roots testing.

The evidence is that the Phillips-Perron tests show a wide range of values for different choices of the lag window and the truncation number in an ARMA (1,1) model. Also, the tests work poorly in a model with heteroscedastic errors. The method of Said-Dickey also has some estimation problems. When the objective function is written in terms of the Kalman Filter, the problem seems to be alleviated. The Lagrange multiplier method also appears unattractive in this case. The deviation is very severe for the Phillips-Perron test compared to others. The normalization of some other information criteria is made, following Ozaki's normalization, for detecting a unit root. For an exact maximum likelihood estimation, the objective function is written such that a switching occurs for a nonstationary or a noninvertible parameter during the estimation. It is found that the Dickey-Fuller statistic $T(\hat{\rho} - 1)$ is sensitive to the process of generating artificial data. In a Monte Carlo study with ARMA (1,1) models, it is also found that the Phillips and Perron tests have greater power than the normalized information criteria only when the truncation point is favorably selected. Other tests seem to have far lower power. As far as unit root testing in an ARMA (1,1) model is concerned, the normalized information criteria seem to be more practical and viable than any other tests.

CHAPTER I
INTRODUCTION

It has been noted that many macroeconomic time series exhibit persistent upward movements, and the pattern of growth in the series resembles a random walk process. With this possible nonstationary behavior, time series data have been suspected to have statistical properties that violate the statistical conditions of constant mean and finite variance postulated for the usual statistical analyses such as estimation, inference, and modelling. Granger and Newbold (1974) showed by a Monte Carlo method that the ordinary least squares regression between the levels of two independent random walk processes produces unreliable conventional test values which may falsely lead to the rejection of the hypothesis of no relation. In the same setting of a general integrated random process, Phillips (1986) developed a rigorous asymptotic theory and proved that many of the Granger-Newbold simulation findings can be explained by an extensive mathematical derivation of the limiting distributions of many commonly-used regression statistics. To avoid the problems occurring from the nonstationary nature, customarily, either a differencing transformation technique suggested by Box and Jenkins (1976) or a polynomial detrending technique used to be employed before the modelling and analysis stage. However, these methods can be problematic too without knowledge of a true data-generating process. Nelson and Kang (1981) report that some undesirable behavior has been

detected for the autocorrelations of the residual obtained by an inappropriate detrending of a random walk process, by the regression of a random walk process on time. Likewise, as Nelson and Plosser (1982) demonstrated, an inaccurate differencing of stationary series around a time trend creates a new series with a noninvertible moving average component. Accordingly, a testing method that can detect the presence of a unit root and the subtle difference between competing representations is very important.

During the last decade, a family of parametric formal testing methods has been developed, which includes Fuller (1976), Dickey and Fuller (1979, 1981), Evans and Savin (1981, 1984), Said and Dickey (1984, 1985), Solo (1984), Phillips (1987), and Phillips and Perron (1986). On the other hand, a family of non-parametric tests has been developed, which includes Campbell and Mankiw (1987a) and Lo and MacKinlay (1989). With these developments, the testing for a unit root in macroeconomic time series has emerged in recent years as a popular subject in econometric literature because the hypothesis of a unit root conveys many important theoretic and empirical implications. Researchers who have employed those testing methods almost unanimously report that most of the macroeconomic time series contain a unit root. An example is the study of historical U.S. macroeconomic time series by Nelson and Plosser (1982). They consider two fundamentally different classes of nonstationary models that compete for a legitimate representation of the secular movements. One class is a deterministic trend-stationary model and the other class is a purely stochastic difference-stationary model. When a unit root is present in the

autoregressive (AR) part of an autoregressive moving average (ARMA) model, a first-differencing is required for inducing stationarity. In this case an innovation has a permanent accumulative effect on the future realization of the variable. In terms of forecasting, the variance of the k-period forecast error for the level of the series will increase without bound as one increases the forecasting period k. On the other hand, when a series is better represented by a stationary fluctuation around a deterministic time trend, an innovation has only a temporary effect, which decays gradually, on the future realization. The long-term forecast error is bounded for the model. They report that most of the macroeconomic time series contain a unit root in the AR part when the unit root testing method of Dickey and Fuller (1979) is employed. Also, they illustrate how an empirical testing result based on the distinction can bring on a different theoretical interpretation about the contribution of a real factor on output variation and macroeconomic fluctuations. Perron (1986) reassesses the findings of Nelson and Plosser (1982) and draws the same conclusion using a newly-developed method. Some other studies are Stock and Watson (1986), Campbell and Mankiw (1987b), and Perron and Phillips (1987) on GNP series; and Meese and Singleton (1982), Corbae and Ouliaris (1986), and Baillie and Bollerslev (1987) on foreign exchange rate data.

The concept of an integrated process that contains unit roots is incorporated in some of the neighboring subjects. The presence of common unit roots, or the same order of integration, in each of a group of variables that constitute a long-run equilibrium relation, forms the basis of the co-integration theory and the error correction dynamic

model suggested in Granger (1981), Granger and Weiss (1983), Engle and Granger (1987), and Hendry (1986). Litterman (1986) obtained improved forecasting results from a vector autoregressive (VAR) model by restricting the coefficients of the first lag term to have a prior mean of one and the standard deviations of the decreasing successive coefficients in a harmonic fashion.

Motivation and Objectives

In recent years, the application of the formal unit root testing methods became very popular, though those methods appear to have some limitations in practical usage. It has been found that many economic times series contain a moving average component when they are first-differenced. This phenomenon is mentioned in Cooper and Nelson (1975), Nelson and Schwert (1977) and Schwert (1987, 1988). For ARMA (1,1) processes, when the AR coefficient is close to or equal to one and the moving average coefficient is very large, say -.8 or -.75, the autocorrelation function of the data shows a very slow decaying pattern with a few positive peaks. Schwert (1987, 1988) presents a remarkable result that when a large moving average component is present in an ARMA (1,1) model, the distribution of the unit roots testing statistics of Said and Dickey (1984, 1985), Phillips (1987), Phillips and Perron (1986), which are designed to handle an ARIMA model, deviate greatly from the empirical distribution constructed by Dickey and Fuller. The result is that many of the tests would falsely reject the unit root hypothesis when the empirical Dickey-Fuller critical values tabulated in Fuller (1976) are used. In practice, those critical values have been

frequently employed without questioning the justifiability of doing that. One implication is that the critical values for the testing should be retabulated according to the length of the sample and the magnitude of the moving average coefficient for each test as Schwert (1988) tried. This is costly and requires tremendous work, though it is not impossible. To make the matters worse, this tabulation cannot be used effectively, because a priori the correct model specification is not known. Accordingly, the usual way of fixing the nominal size of the Type I error at α, say .05 or .01, becomes less meaningful. It is important in this situation to choose a unit root testing method, among several frequently used, that is more dependable than others.

The objective of this study is to investigate the problems with each of the conventional unit roots testing methods, and to evaluate their power in useful specific conditions in assessing an autoregressive integrated moving average model (ARIMA) of order (0,1,1) and an ARIMA (1,0,1) model, or ARMA (1,1), when there exists a large moving average coefficient approaching the autoregressive coefficient. It will be investigated how suitable the normalized information criteria would be, which rarely got much attention. Finally, better method from the practical point of view will be chosen.

In this study, the methods considered are based on Phillips and Perron (1986), Said and Dickey (1985), Solo (1984), and the normalized Akaike's information criteria (hereafter the normalized AIC) of Ozaki (1977) and two extensions thereof. The first three methods have a common feature that they converge to the limiting distribution of the Dickey-Fuller tests. In this study, they are called the Dickey-Fuller

line of unit roots testing. The normalization method of Ozaki (1977) is similarly extended to the criteria of Schwarz (1978) and Hannan and Quinn (1979). The three criteria constitute the normalized information criteria.

All these methods can serve the same purpose, while they have different characteristics. The tests of Phillips and Perron (hereafter called Phillips-Perron tests) are the most general in that they can handle the model with a more general assumption about the statistical nature of the error term. These tests are most commonly used because of theoretical elegance and ease of use. The tests of Said and Dickey, Solo, and Ozaki are designed to test the order of integration in an ARIMA model with more stringent assumptions on the error term. These tests are harder to implement than the Phillips-Perron tests because they rely on nonlinear estimation instead of ordinary least squares and subsequent nonparametric adjustment. Solo's Lagrange Multiplier tests (hereafter LM tests) need to be evaluated only under the null hypothesis that unit roots are present, contrary to the Said-Dickey tests which need to be evaluated only under the alternative. The normalized information criteria need evaluation under both hypotheses. These diversities will create a different outcome for each test.

Eventually, this study will contribute to a better understanding of the practicality of unit roots testing methods and their behavior when an autoregressive coefficient is close to the boundary of stationarity and a moving average coefficient is very large, which is believed to occur in many macroeconomic time series.

Scope of the Study and Overview

The scope of the power study is limited to the testing for the presence of a unit root in an ARIMA (1,0,1) model and an ARMA (1,1) model, especially when there exists a large moving average coefficient and, consequently, the series approach a white noise process. As demonstrated by Schwert (1987), many macroeconomic time series fall within this scope. During the simulation experiments, the length of data is fixed at 150 which roughly corresponds to the length of quarterly series after World War II. Because of limited computational time available, the simulation study is performed on a small scale. For most of the study, no ready-to-use econometric software packages are available. All the programs needed for the analysis had to be programmed and fully tested. These considerations led to a restricted simulation study but the results obtained are very useful in practice.

Chapter II investigates a few problematic features of the Phillips-Perron tests, Said-Dickey tests, and LM tests, which are observed during the implementation of the tests. It also discusses the theoretical background of these tests. Chapter III presents Ozaki's normalized version of AIC for order selection in a general ARIMA (p,d,q) model. Extensions to similar criteria such as the Bayesian criterion of Schwarz (SBC) and the order selection rule of Hannan and Quinn (HQ) are made. The performance of each test for selecting the correct order of integration is compared. An application of these criteria to the problem of distinguishing between the trend-stationary (TS) and difference-stationary (DS) models is also discussed. Chapter V presents the computational details of the simulation and the estimation. The

Monte Carlo study is pursued for every testing method for the testing of a unit root in the two ARMA models, in which $(\rho,\theta) = (1,-.8)$ for one model and $(\rho,\theta) = (.95,-.8)$ for the other. A power comparison is made. In chapter V, the findings obtained in this study are summarized and appropriate conclusions are drawn.

CHAPTER II
THE DICKEY-FULLER LINE OF TESTING FOR UNIT ROOTS

The formal testing of a unit root in autoregressive time series models was initiated by Dickey and Fuller during the 1970s and early 1980s. Since their work has provided the basis for later developments of formal testing, some of the main points of their approach are reviewed here.

The equations considered are the three AR (1) type equations defined as follows:

$$y_t = \rho y_{t-1} + u_t, \tag{2.1}$$

$$y_t = \mu + \rho_\mu y_{t-1} + u_t, \tag{2.2}$$

$$y_t = \mu + \beta[t-(T+1)/2] + \rho_\tau y_{t-1} + u_t, \quad t = 1, 2, \ldots, T, \tag{2.3}$$

where $y_o = c$ is a fixed constant and $\{u_t\}$ is a sequence of normal independent random variables with mean 0 and variance σ^2, $NI(0,\sigma^2)$. Among the ten testing statistics derived from equation (2.1) to (2.3), only the ones frequently used will be discussed. Under the null hypothesis of $\rho = 1$, $\mu = 0$, and $c = 0$, they show that the normalized statistics of the least squares estimate for the slope, $T(\hat{\rho} - 1)$ and $T(\hat{\rho}_\mu - 1)$, and t-statistics for the testing of the null hypothesis, $\hat{\tau}$ and $\hat{\tau}_\mu$, converge respectively to distributions which are some functions

of an integral defined on a Wiener process. Likewise, under the null hypothesis of $\rho_\tau = 1$ and $\beta = 0$, irrespective of the value of μ, $(\hat{\rho}_\tau - 1)$ and $\hat{\tau}_\tau$ converge to distributions which are some function of an integral defined on a Wiener process. If μ is not zero in (2.2) or β is not zero in (2.3), then the limiting distributions of the statistics become normally distributed, but depend upon the unknown parameters. Likelihood ratio statistics such as Φ_1 of the null hypothesis $(\mu,\beta,\rho) = (0,0,1)$ versus the alternative hypothesis $(\mu,\beta,\rho) = (\mu,0,\rho)$, Φ_2 of $(\mu,\beta,\rho) = (0,0,1)$ versus $(\mu,\beta,\rho) = (\mu,\beta,\rho)$, and Φ_3 of $(\mu,\beta,\rho) = (\mu,0,1)$ versus $(\mu,\beta,\rho) = (\mu,\beta,\rho)$ are also shown to have limiting distributions which do not depend on μ and β. Because of this independence from other parameters, Dickey and Fuller could construct empirical distribution tables by a Monte Carlo method. These tables are available in Fuller (1976) and Dickey and Fuller (1981). Dickey and Fuller extended the unit root testing to higher order autoregressive models and demonstrated that the testing statistics suggested for AR(1) models, with a slight correction if necessary, share the same corresponding limiting distributions as in AR(1) models. Therefore, in large samples the same critical values from the tables can be used. A general autoregressive model AR(p) is represented by

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + u_t, \tag{2.4}$$

where the characteristic equation is given by

$$1 - \alpha_1 L - \ldots - \alpha_p L^p = 0. \tag{2.5}$$

If a unit root is suspected and the other p - 1 roots lie outside the unit circle in (2.5), then equation (2.4) can be legitimately written, in order to isolate the unit root on the first coefficient, as

$$y_t = \rho y_{t-1} + \sum_{i=1}^{(p-1)} \rho_i \nabla y_{t-i} + u_t. \tag{2.6}$$

When a unit root exists , the relations $\rho_i = \sum_{j=i}^{(p-1)} \alpha_j$, for $i = 1, 2, \ldots, p-1$ and $\rho = \sum_{i=1}^{p} \alpha_i$ hold. Similarly, using the same reasoning, equations (2.2) and (2.3) are shown to have the following counterparts,

$$y_t = \mu + \rho_\mu y_{t-1} + \sum_{i=1}^{(p-1)} \rho_i \nabla y_{t-i} + u_t, \tag{2.7}$$

$$y_t = \mu + \beta[t-(T+1)/2] + \rho_\tau y_{t-1} + \sum_{i=1}^{(p-1)} \rho_i \nabla y_{t-i} + u_t. \tag{2.8}$$

As in the previous case, ordinary least squares estimation is applied to the equations to produce related quantities for testing.

Suppose that an ARIMA(p,1,q) process, where the error terms u_t are invertible moving average processes, is assumed. Then the testing equations are given by

$$y_t = \rho y_{t-1} + \sum_{i=1}^{(p-1)} \rho_i \nabla y_{t-i} + u_t + \sum_{i=1}^{q} \theta_j u_{t-j}, \tag{2.9}$$

$$y_t = \mu + \rho_\mu y_{t-1} + \sum_{i=1}^{(p-1)} \rho_i \nabla y_{t-i} + u_t + \sum_{i=1}^{q} \theta_j u_{t-j}, \tag{2.10}$$

$$y_t = \mu + \beta[t-(T+1)/2] + \rho_\mu y_{t-1} + \sum_{i=1}^{(p-1)} \rho_i \nabla y_{t-i} + u_t$$
$$+ \sum_{i=1}^{q} \theta_j u_{t-j}. \tag{2.11}$$

The Dickey-Fuller tests which are valid up to a higher order autoregression are not valid for the above mixed ARMA models.

Accordingly, an extension to more general models than autoregressive models became inevitable.

The Dickey-Fuller tests have been extended in two directions. The bifurcation may be related to the estimation method employed in a specific testing method. In one direction, they stick to an ordinary least squares regression. Said and Dickey (1984) tests and Phillips-Perron tests seem to follow this line. In the other direction, they employ a nonlinear estimation method as in Said and Dickey (1985) and the LM tests of Solo (1984). The approach of Said and Dickey (1984), which approximates an ARIMA(p,1,q) by a high order AR(p^*), where p^* is a function of the number of observations, will not be pursued in this study. Some problematic features of each of the remaining three methods will be investigated in this study.

The contribution of Dickey and Fuller is that their studies initiated a formal testing of unit roots in the setting of an autoregressive model and stimulated future research on this topic. Dickey and Fuller investigated the limiting distributions of the relevant statistics using ordinary least squares, and tabulated the empirical distribution of each of these tests.

The Phillips-Perron Tests

So far, the Phillips-Perron tests can be viewed as the most general extensions of the Dickey-Fuller tests, as the assumptions made about the innovation sequence u_t in (2.1), (2.2) and (2.3) are very general. The assumptions allow u_t to have some degree of temporal dependence and heteroscedasticity in the process such that it can represent not only

stationary and invertible ARMA models of unknown order, but also autoregressive conditional heteroscedasticity (ARCH) models. In the tests, equations (2.1), (2.2) and (2,3) are estimated by the least squares estimation method, and residuals are obtained. Next, the usual Dickey-Fuller statistics are calculated and the appropriate non-parametric adjustment terms, constructed with the residuals, are subtracted from each statistic. These new tests are shown to have the same limiting distributions as those of the Dickey-Fuller tests found in earlier work under strict assumptions on u_t.

The idea behind the extension is that when a unit root is present, a realization of y variable at time t in equations (2.1), (2.2), and (2.3) can be expressed, by a repetitive substitution, in terms of a partial sum, S_t, of innovations stemming from time zero up to t and the initial value. By the result of a functional central limit theorem, a standardized partial sum, Z_t, of a wide range of nonstationary, weakly dependent, and heterogeneously distributed innovation sequences as

$$Z_t(r) = T^{1/2}S_{r-1}/\sigma, \qquad (j - 1)/T \le r < j/T, \quad (j = 1, 2, \ldots, T),$$

$$Z_t(1) = T^{1/2}S_T/\sigma,$$

$$\sigma^2 = \lim_{T\to\infty} E(T^{-1}S_T^2),$$

can be shown to converge weakly to a Wiener process under very general conditions. By employing the continuous mapping theorem, all the tests suggested by Dickey and Fuller can be shown to have limiting distributions that are functional of a Wiener process. When a general

error process is assumed, the limiting distributions of $(\hat{\rho} - 1)$ and $\hat{\tau}$ of model (2.1) under the null hypothesis are

$$(\hat{\rho} - 1) \Rightarrow (1/2)(W(1)^2 - \sigma_u^2/\sigma^2) \Big/ \int_0^1 W(r)^2\, dr$$

and

$$\hat{\tau} \Rightarrow (\sigma/2\sigma_u)(W(1)^2 - \sigma_u^2/\sigma^2) \Big/ \left\{ \int_0^1 W(r)^2\, dr \right\}^{1/2},$$

where $\sigma^2 = \lim_{T\to\infty} E(T^{-1}S_T^2)$ and $\sigma_u^2 = \lim_{T\to\infty} T^{-1}E(u_T^2)$, and $W(r)$ is a standard Wiener process on the space of all real valued continuous functions defined on [0,1]. The earlier studies of Dickey and Fuller considered the special case when $\sigma_u^2 = \sigma^2$. Accordingly, if σ_u^2 and σ^2 can be estimated consistently and some manipulation is made over the Dickey and Fuller tests such that the ratio σ_u^2/σ^2 becomes one in a large sample, then these new tests can have the same limiting distributions as those of the Dickey and Fuller statistics. Since $\hat{\rho}$ converges to one under the null hypothesis of a unit root, the variance of the error term σ_u^2 is easily estimated by s_u^2,

$$s_t^2 = T^{-1} \sum_{t=1}^{T} (y_t - \hat{\rho} y_{t-1})^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2.$$

But, the estimation of the variance of the partial sum, σ^2, is not easy. Phillips (1987) recommends a weighted variance estimator as follows:

$$s_{TL}^2 = T^{-1} \sum_{t=1}^{T} \hat{u}^2 + 2T^{-1} \sum_{\tau=1}^{L} \omega_{\tau L} \sum_{t=\tau+1}^{T} \hat{u}_t \hat{u}_{t-\tau} \tag{2.12}$$

where

$$\omega_{\tau L} = 1 - \tau/(L+1).$$

The new tests for unit roots obtained by adjusting $T(\hat{\rho}-1)$ and $\hat{\tau}$ in the equations (2.1), (2.2), and (2.3) are as follows,

$$Z_{\rho} = T(\hat{\rho} - 1) - 0.5(s_{TL}^2 - s_u^2)\cdot(T^{-2}\sum_{t=1}^{T} y_{t-1}^2)^{-1}, \tag{2.13}$$

$$Z_{\tau} = \hat{\tau}(s_u/s_{TL}) - 0.5(s_{TL}^2 - s_u^2)\cdot\left\{ s_{TL}^2 T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \right\}^{-1/2}, \tag{2.14}$$

$$Z_{\rho\mu} = T(\hat{\rho}_{\mu} - 1) - 0.5(s_{TL}^2 - s_u^2)\cdot[T^{-2}\sum_{t=1}^{T}(y_{t-1} - \bar{Y})^2]^{-1}, \tag{2.15}$$

$$Z_{\tau\mu} = \hat{\tau}_{\mu}(s_u/s_{TL}) - 0.5(s_{TL}^2 - s_u^2)\cdot\left\{ s_{TL}^2 T^{-2}\sum_{t=1}^{T}(y_{t-1} - \bar{y})^2\right\}^{-1/2}, \tag{2.16}$$

$$Z_{\rho\tau} = T(\hat{\rho}_{\tau} - 1) - (s_{TL}^2 - s_u^2)\cdot\{T^6/[24\cdot\det(X'X)]\}, \tag{2.17}$$

and

$$Z_{\tau\tau} = \hat{\tau}_{\tau}(s_u/s_{TL}) - 0.5(s_{TL}^2 - s_u^2)\cdot T^3\{4s_{TL}[3\cdot\det(X'X)]^{1/2}\}^{-1} \tag{2.18}$$

respectively, where $X'X$ is the moment matrix of appropriate right hand side variables. Many other testing statistics are also extensively tabulated in the appendix of Perron (1986).

The Phillips-Perron tests have been applied popularly to many macroeconomic time series because of their theoretical elegance and general applicability. Theoretically, the tests can be used with any time series considered in economics in normal situations, without a detailed knowledge about the behavior of innovations and model identification. Despite this merit, it is noted that in practice the tests always carry some fundamental ambiguities in estimating the variance of the partial sum σ^2.

The weighted variance estimator (2.12) contains a set of weights originating from a non-negative spectral density function, which corresponds to the Bartlett lag window to insure the positive definiteness of the variance. Here, the problem of "window carpentry" emerges. There are several other equally eligible choices of lag window which also provide consistent estimates of spectral density functions. Another problem is that there exists no optimal rule for selecting an adequate truncation point.

The problems of the Phillips-Perron tests are illustrated by examining the data pertaining to 197 concentration readings from a chemical process listed in Box and Jenkins (1976, p. 525). Box and Jenkins fitted two competing models:

$$z_t = 1.45 + \underset{(.04)}{.92} z_{t-1} + \hat{u}_t - \underset{(.08)}{.58} \hat{u}_{t-1} \tag{2.19}$$

and

$$\nabla z_t = \hat{u}_t - \underset{(.05)}{.70} \hat{u}_{t-1}. \tag{2.20}$$

With the same data, Phillips and Perron (1986) fitted the following equation by ordinary least squares,

$$y_t = \underset{(1.002)}{7.300} + \underset{(.059)}{.572} y_{t-1} + \hat{u}_t. \tag{2.21}$$

They choose the lag truncation point of L = 1, since they believe that, as in any ARIMA (0,1,1) model and ARMA (1,1) model, the errors are well represented by a first-order moving average. They obtained -6.973 and -74.201 for $Z_{\tau\mu}$ and $Z_{\rho\mu}$. Accordingly, they rejected the null hypothesis

Table 2-1
Autocorrelation of residuals for the 197 chemical process concentration readings from Box and Jenkins (1976).

	residuals from			
lag	(2.2)	(2.3)	(2.19)	(2.20)
1	-.146	-.146	-.335	-.415
2	.151	.151	.071	.019
3	.063	.064	-.012	-.067
4	.092	.094	.044	-.011
5	.044	.045	-.014	-.070
6	.086	.088	.034	-.021
7	.220	.222	.200	.151
8	.041	.043	-.011	-.071
9	.116	.118	.092	.039
10	.089	.092	.075	.022
11	.020	.022	.010	-.049
12	-.002	.000	-.008	-.068
13	.048	.050	.045	-.012
14	.019	.020	.219	.172
15	-.089	-.088	-.121	-.185
16	.082	.083	.089	.036
17	.071	.073	.070	.015
18	.127	.129	.142	.090
19	-.057	-.055	-.074	-.136
20	.177	.180	.215	.168

of ARIMA (0,1,1) against ARMA(1,1) using the tabulated critical values of Dickey and Fuller. But, their arguments are not sound. The same data are analyzed, and it is found that the autocorrelation function constructed by the least squares residuals in (2.21) shows no clear pattern that leads them to pick first-order moving average as the best representation. Table 2-1 lists the autocorrelations of the residuals obtained by estimating the four different equations given by (2.2), (2.3), (2.19), and (2.20). The autocorrelations constructed from the residuals of ARMA (1,1) or ARIMA (0,1,1) models, presented in the third and fourth columns of the table, clearly show a quick drop after the first lag like that of a typical first-order moving average process. On the contrary, the autocorrelations from the least squares residuals of fitting (2.2) or (2.3), presented in the first and second columns of the table, show a pattern which is hard to interpret. The autocorrelations do not damp quickly after the first lag and the peak at the seventh lag appears to be large. As is shown for the chemical data, the choice of truncation point in the Phillips-Perron tests is not readily seen. Accordingly, it may be doubtful whether Phillips-Perron type nonparametric corrections will work and whether the tests can effectively distinguish between (2.20) and (2.21) by choosing L = 1 or some other value of L. Other window options are investigated in the following. As the sample mean of the data is 17.06, which is not close to zero, the choice of Phillips-Perron test statistics $Z_{\rho\tau}$ and $Z_{\tau\tau}$ are the first priority of interest, and $Z_{\rho\mu}$ and $Z_{\tau\mu}$ next. Phillips suggested employing the Bartlett window, while there is a report that the properties of the Bartlett window are inferior to those of the Tukey

and Parzen windows. (Refer to Chatfield [1984], p. 141.) Accordingly, together with the Bartlett lag window, two additional lag windows are considered: the Tukey window

$$w_{\tau L} = .5[1+\cos(\pi\tau/L)], \qquad \tau = 1,2,\ldots,L,$$

and the Parzen window

$$w_{\tau L} = \begin{cases} 1 - 6(\tau/L)^2 + 6(\tau/L)^3, & 1 \le \tau < .5L, \\ 2(1-\tau/L)^3, & .5L \le \tau \le L, \end{cases}$$

for different values of the truncation point L. Table 2-2 contains t-statistic values of the Phillips-Perron tests when equations (2.2) and (2.3) incorporating different windows and truncation points are estimated. The tests using the Bartlett window shows ever decreasing values as L is increased. And the tests incorporating either the Parzen window or the Tukey window reach a maximum around L = 3 and thereafter exhibit a decreasing pattern. For all equations, the statistics can have a wide range of values depending largely on the choice of truncation point L and, to a lesser degree, the choice of lag window. The normalized first-order autocorrelation coefficient of the Phillips-Perron tests is listed in table 2-3. The patterns are quite similar to the those of the t-statistics except that the variations in the quantities are much more intensified. Without an optimal rule for the choice of the truncation point, there may exist room for arbitrariness. Selecting either the most favorable or the least favorable values of the statistics can be equally possible. Phillips and Perron mistakenly used the critical value constructed by Dickey and Fuller, which is only

Table 2-2
Variations of the t-statistics of the Phillips-Perron tests for a unit root applied to the 197 chemical process concentration readings from Box and Jenkins (1976) according to three lag windows and truncation points.

	Bartlett		Parzen		Tukey	
lag	$Z_{\tau\tau}$	$Z_{\tau\mu}$	$Z_{\tau\tau}$	$Z_{\tau\mu}$	$Z_{\tau\tau}$	$Z_{\tau\mu}$
1	-6.945	-6.909	-7.196	-7.224	-7.195	-7.224
2	-7.042	-7.020	-7.029	-7.067	-6.923	-6.909
3	-7.125	-7.143	-6.900	-6.922	-6.893	-6.913
4	-7.278	-7.293	-6.921	-6.941	-7.031	-7.050
5	-7.378	-7.423	-7.045	-7.031	-7.200	-7.208
6	-7.495	-7.565	-7.138	-7.139	-7.331	-7.364
8	-7.956	-7.961	-7.355	-7.368	-7.676	-7.685
10	-8.217	-8.327	-7.563	-7.620	-7.940	-8.062
12	-8.616	-8.608	-7.884	-7.889	-8.434	-8.431
14	-8.776	-8.875	-8.066	-8.157	-8.620	-8.746
16	-9.112	-9.094	-8.413	-8.410	-9.037	-9.023
18	-9.224	-9.322	-8.550	-8.643	-9.157	-9.267
20	-9.561	-9.531	-8.869	-8.857	-9.518	-9.493
22	-9.644	-9.716	-8.971	-9.055	-9.611	-9.704
24	-9.923	-9.883	-9.258	-9.238	-9.934	-9.898
26	-9.983	-10.044	-9.337	-9.408	-10.005	-10.077
28	-10.253	-10.203	-9.596	-9.567	-10.291	-10.245
30	-10.308	-10.343	-9.660	-9.718	-10.351	-10.404

Note: $Z_{\tau\tau}$ and $Z_{\tau\mu}$ are defined in (2.18) and (2.16) respectively.

Table 2-3
Variations of the normalized first-order autocorrelation coefficient of the Phillips-Perron tests applied to the 197 chemical process concentration readings from Box and Jenkins (1976) according to three lag windows and truncation points.

	Bartlett		Parzen		Tukey	
lag	$Z_{\rho\tau}$	$Z_{\rho\mu}$	$Z_{\rho\tau}$	$Z_{\rho\mu}$	$Z_{\rho\tau}$	$Z_{\rho\mu}$
1	-73.33	-73.52	-82.93	-83.15	-82.93	-83.15
2	-76.76	-76.91	-78.13	-78.34	-73.33	-73.52
3	-80.55	-80.66	-73.73	-73.92	-73.50	-73.65
4	-85.25	-85.31	-74.36	-74.51	-77.70	-77.81
5	-89.36	-89.37	-77.10	-77.23	-82.61	-82.67
6	-93.91	-93.85	-80.47	-80.56	-87.51	-87.52
8	-106.88	-106.66	-87.64	-87.65	-97.79	-97.67
10	-119.39	-118.99	-95.70	-95.62	-110.28	-110.02
12	-129.35	-128.75	-104.50	-104.30	-123.00	-122.56
14	-139.10	-138.28	-113.53	-113.20	-134.28	-133.63
16	-147.30	-146.28	-122.29	-121.83	-144.53	-143.68
18	-156.06	-154.83	-130.59	-129.98	-153.82	-152.75
20	-164.29	-162.83	-138.40	-137.63	-162.63	-161.35
22	-171.71	-170.02	-145.76	-144.84	-171.07	-169.56
24	-178.57	-176.65	-152.74	-151.65	-179.02	-177.28
26	-185.30	-183.16	-159.35	-158.10	-186.48	-184.51
28	-192.03	-189.67	-165.66	-164.25	-193.59	-191.40
30	-190.07	-195.48	-171.71	-170.13	-200.45	-198.03

Note: $Z_{\rho\tau}$ and $Z_{\rho\mu}$ are defined in (2.17) and (2.15) respectively.

asymptotically justifiable. But, those critical values should not be used with a sample of length 200. A Monte Carlo study of Schwert (1988) reveals that, in a mixed ARIMA (0,1,1) model, the Phillips-Perron testing statistics, even with samples of length as large as 10,000, do have different finite sample distributions from the ones suggested by Fuller (1976) and Dickey and Fuller (1979, 1981) for autoregressive processes so that the null hypothesis is too often falsely rejected. Accordingly, in order to apply the Phillips-Perron tests, accurate critical values need to be found by some means, say by a large simulation, for each case.

Phillips and Perron show elegantly the limiting distributions of many unit root testing statistics and claim that their versions can be applied to very general cases. But it appears that theoretical elegance and practicality do not go hand in hand. According to Schwert (1987), the Phillips-Perron test statistics converge extremely slowly to the limiting distributions when a moving average component is present. The employment of the Phillips-Perron tests without the knowledge of correct specification can lead to a false conclusion. With the help of conventional ARIMA fitting procedures, identification, estimation and diagnostic checking, some knowledge of the specification may be acquired. Even with the information obtained, the critical values for the testing in a finite sample are not readily available. Those values for each statistic are known to vary widely according to the magnitude of moving average coefficients, sample length, different truncation points L, and also the choice of the lag window as shown previously. The problem associated with the variability is that the conventional way

of fixing the test size at a certain level becomes very difficult. Additionally, an extensive work for the tabulation is needed. A limited tabulation is found in Schwert (1987).

Another interest is how the Phillips-Perron tests will perform in a more general situation. In the following, the random walk series without a drift, of which the errors are temporarily dependent and heteroscedastic, are constructed. The model considered is

$$y_t = y_{t-1} + u_t, \qquad t = 1, 2, \ldots, T,$$

where u_t is equal to $\sigma_t\lambda_t$, the sequence $\{\lambda_t\}$ is a sequence of NI(0,1) and σ_t has the specific form of heteroscedasticity of

$$\ln \sigma_t^2 = \psi \ln \sigma_{t-1}^2 + \xi_t,$$

where the sequence $\{\xi_t\}$ is a sequence of NI(0,1). Conditional upon σ_t^2, u_t is normally distributed with mean zero and variance σ_t^2 and λ_t and ξ_t are assumed to be independent. These series are considered in Lo and Mackinlay (1989), French, Schwert, and Stambaugh (1987), and Porteba and Summers (1986). Table 2-4 contains the empirical sizes of Phillips and Perron testing statistics $Z_{\tau\mu}$ and $Z_{\rho\mu}$ applied to the random walk series defined earlier. As the ψ coefficient, which is assumed to take any value between zero and one, approaches one from $\psi = .5$, the Phillips-Perron tests based on both t-statistics and normalized first-order autocorrelation coefficient begin to deviate from the distribution of the Dickey-Fuller tests in samples of length 150. When the degree of heterogeneity and dependence of the process is high, the size of the

Table 2-4
Empirical sizes of the Phillips-Perron tests for unit roots based on Dickey-Fuller tabulation when the true model is a random walk with heteroscedastic disturbances.

ψ value	D-F test size	$Z_{\tau\mu}$		$Z_{\rho\mu}$	
		L = 4	L = 12	L = 4	L = 12
.95	10.0 %	23.8%	24.3%	23.3%	23.7%
	5.0 %	17.3%	17.7%	18.7%	18.8%
	2.5 %	13.9%	14.0%	13.9%	14.3%
	1.0 %	10.5%	11.3%	9.9%	10.2%
.9	10.0 %	18.6%	19.5%	17.0%	18.0%
	5.0 %	12.4%	12.5%	12.7%	13.7%
	2.5 %	9.0%	9.5%	8.5%	9.5%
	1.0 %	6.7%	6.6%	5.5%	6.4%
.7	10.0 %	12.2%	12.9%	11.1%	11.8%
	5.0 %	7.7%	8.0%	7.1%	7.5%
	2.5 %	4.5%	4.6%	4.6%	4.9%
	1.0 %	1.8%	2.0%	1.9%	2.4%
.5	10.0 %	10.8%	11.5%	9.1%	10.9%
	5.0 %	5.7%	6.5%	6.1%	6.7%
	2.5 %	3.4%	3.4%	3.2%	3.6%
	1.0 %	1.6%	1.8%	1.4%	1.5%

<u>Note</u>: $Z_{\tau\mu}$ and $Z_{\rho\mu}$ are defined in (2.16) and (2.15) respectively. Based on 2,000 replications of a process, consists of sample length 150,

$$y_t = y_{t-1} + u_t, \qquad t = 1, 2, \ldots, T.$$

where u_t equals to $\sigma_t\lambda_t$, $\{\lambda_t\}$ is a sequence of NI(0,1), and σ_t has the specific form of heteroscedasticity of

$$\ln \sigma_t^2 = \psi \ln \sigma_{t-1}^2 + \xi_t,$$

where $\{\xi_t\}$ is a sequence of NI(0,1). λ_t and ξ_t are assumed to be independent. The first 50 observations have been discarded for each replication to eliminate the start-up effects.

tests becomes very sensitive to a small change of the ψ coefficient. As in the mixed ARIMA models with normal errors, for the models with heteroscedastic and temporarily dependent errors, the correct model specification and the corresponding critical values must be known before the Phillips-Perron tests are applied. In the circumstances, obtaining a consistent estimate of the ψ coefficient does not guarantee accurate testing, and, therefore, using the critical values of Dickey (1976) can be even more inadequate and misleading.

For the process with a more general error structure than normal error, it is difficult to know the correct specification with accurate parameters and also the finite sample behavior of the Phillips-Perron tests for the process. At this state of knowledge, from the practical point of view, the effectiveness of the Phillips-Perron tests in evaluating a finite sample process with a more general error structure is significantly limited. It naturally follows that the Phillips-Perron tests do not dominate other methods of testing for unit roots.

The Said-Dickey Tests

Said and Dickey (1985) proposed a more general unit root testing method in ARMA (p,q) models, compared with Dickey-Fuller, in which they directly estimate moving average coefficients nonlinearly. During the derivation of this testing method, they rely on the estimation procedure from Fuller (1976), which is a one-step Gauss-Newton nonlinear estimation. It is shown that the idea can be illustrated with ease within a simple ARMA (1,1) model under the null, and the results can be extended to general ARMA (p,q) models easily. They consider the time series satisfying

$$y_t = \rho y_{t-1} + u_t + \theta u_{t-1}, \qquad t = 1, 2, \ldots, T, \tag{2.22}$$

where $y_o = 0$, $|\theta| < 1$, and $\{u_t\}$ is a sequence of normal independent random variables with zero mean and a finite variance. Under the initial condition $u_o = \delta$ and the null hypothesis $\rho = 1$, an ARMA (1,1) model can be expressed by a repetitive substitution as

$$u_t = y_t - \sum_{i=0}^{t-2}(-\theta)^i(\rho + \theta)y_{t-i-1} + (-\theta)^t\delta. \tag{2.23}$$

By a Taylor series expansion of (2.23) about the estimated coefficient vector $\hat{\Phi}' = (\hat{\rho}|\hat{\theta}|\hat{\delta})$, $u_t(\Phi)$ is given by

$$u_t(\Phi) = u_t(\hat{\Phi}) - V_t(\hat{\Phi})(\rho - \hat{\rho}) - W_t(\hat{\Phi})(\theta - \hat{\theta})$$

$$- \Delta_t(\hat{\Phi})(\delta - \hat{\delta}) + r_t, \tag{2.24}$$

where $-V_t(\hat{\Phi})$, $-W_t(\hat{\Phi})$, and $-\Delta_t(\hat{\Phi})$ are the partial derivatives of $u_t(\Phi)$ with respect to ρ, θ, and δ evaluated at $\hat{\Phi}$, and r_t represents the remainder term. As the remainder can be ignored under the null hypothesis and $u_t(\Phi)$ are distributed as $NI(0,\sigma^2)$, they suggest regressing $u_t(\hat{\Phi})$ on $V_t(\hat{\Phi})$, $W_t(\hat{\Phi})$, and $\Delta_t(\hat{\Phi})$ to get an improved estimator of the true parameter vector Φ starting with consistent estimates in an iterative scheme. Said and Dickey show that the normalized coefficient statistics $T(\hat{\rho} - 1)$ and $\hat{\tau}$, $T(\hat{\rho}_\mu - 1)$ and $\hat{\tau}_\mu$, when a nonzero mean is removed, respectively share the same limiting distributions as the corresponding distributions of Dickey and Fuller (1979). For an easy implementation, an ARMA (p,q) model is written as follows:

$$u_t = Z_t + \alpha_1 Z_{t-1} + \cdots + \alpha_{p-1} Z_{t-p+1} +$$

$$\theta_1 u_{t-1} + \cdots + \theta_q u_{t-q}, \qquad (2.25)$$

where $Z_t = y_t - \rho y_{t-1}$. The procedure taken for an ARMA (1,1) model is similarly followed for ARMA (p,q) models. But, additional derivatives should be included as right hand side variables as follows:

$$u_t(\hat{\Phi}) = V_t(\hat{\Phi})(\rho - \hat{\rho}) + \sum_{i=1}^{p-1} X_{it}(\hat{\Phi})(\alpha_i - \hat{\alpha}_i)$$

$$+ \sum_{j=1}^{q} W_{jt}(\hat{\Phi})(\theta_j - \hat{\theta}_j) + \sum_{k=1}^{q} \Delta_{kt}(\hat{\Phi})(\delta_k - \hat{\delta}_k) + u_t.$$

For general ARMA (p,q) models, Said and Dickey suggest using the testing statistics $T(\hat{\rho} - 1)$, $\hat{\tau}$, $T(\hat{\rho}_\mu - 1)$, and $\hat{\tau}_\mu$ considered in the ARMA (1,1) model since their limiting distributions do not change.

One aspect of the Said-Dickey method is that the asymptotic properties of the testing statistics do not change whether a single estimation step is executed or an iterative estimation method is used. Said and Dickey (1985) revealed, by a simulation study, that, when the one-step Gaussian estimation method is employed, the power is highly affected by the starting value of the moving average coefficient in ARMA (1,1) models. When the true coefficient was used as the starting value, the size of the test appeared to be quite close to the nominal level. But, in practice, the true parameter value is not available and a consistent estimate is desired as an initial value. Unfortunately, in the method, there is no criterion to choose a specific consistent initial estimate among possible competing consistent estimates. Initial estimates that are consistent under both the null and alternative

hypotheses were tried by them, but the result is not satisfactory. Here, the iterative estimation method is employed, as the results may be less sensitively affected by the choice of consistent initial estimates than in the previous case. Throughout the iterative estimation, $\hat{\rho}$ is restricted to 1, and the other updated estimates after each iteration are used as initial values for the next iteration.

It is noted that employing equation (2.23) suggested by Said and Dickey as the objective function for evaluating residuals may be less attractive when the moving average coefficient is large. A simulation study reveals that when the moving average coefficient θ, fixed at -.8 in the simulation, is close to -1, in a mixed ARIMA (0,1,1) process, about 6 percent of the time the coefficients did not converge in a reasonably large number of iterations, which in this study is 25. In thirty two cases out of five hundred replications, with many different starting values tried for each estimation of the coefficients, the iterations did not converge when the convergence criterion .00002 was used. For those cases, the coefficients bounced back and forth without achieving a convergence, or the speed of the convergence was extremely slow. To investigate a possibility to alleviate the problem, the objective function suggested by Said and Dickey was replaced by the one expressed in terms of the Kalman algorithm. (Refer to Harvey and Phillips [1979] for details.) The initial condition of $y_0 = 0$ is incorporated and $u_0 = \delta$ is included in the starting error vector as an estimable parameter. In seventeen cases out of five hundred replications, the coefficients did not converge and exhibited the pattern already noted. Table 2-5 lists the testing statistics values

Table 2-5
A comparison of coefficient estimates and the unit roots testing statistics obtained by a one-step Gauss-Newton nonlinear estimation using the objective function suggested by Said and Dickey (1985) and the objective function based on Kalman algorithm for the mixed ARIMA (0,1,1) processes, where the moving average coefficient θ=-.8.

Said and Dickey			Kalman algorithm		
$T(\hat{\rho} - 1)$	$\hat{\tau}$	$\hat{\theta}$	$T(\hat{\rho} - 1)$	$\hat{\tau}$	$\hat{\theta}$
-2.760	-4.290	-1.033 (-5422)	-1.096	-1.744	-.930 (-30)
-.134	-.455	-1.041 (-742)	.001	.172	-.993 (-22)
1.156	1.875	-1.039 (-1296)	.625	.638	-.904 (-21)
.635	1.589	-1.038 (-1575)	.355	1.169	-.973 (-27)
2.935	1.080	-1.011 (-67)	1.241	.887	-.878 (-18)
.381	1.481	-1.045 (-3298)	1.000	.556	-.959 (-30)
-.664	-.786	-1.035 (-627)	.213	.205	-.912 (-24)

<u>Note</u>: Both of the statistics $T(\hat{\rho} - 1)$ and $\hat{\tau}$ are obtained by employing the regression (2.24). The numbers in the above parentheses represent the standard errors associated with each moving average coefficient. During the simulation, the first 50 observations have been discarded to eliminate the start-up effect.

calculated by the last iteration values of the coefficients for the six cases, where the estimation method of Said and Dickey could not produce an invertible moving average coefficient estimate, and the comparable converged values obtained employing the objective function in terms of

Kalman filter algorithm. The conventional t-test of the moving average coefficient based on the Said-Dickey method seems to be greatly exaggerated compared with that based on the latter method when the coefficients do not have an invertible estimate. An interesting result is that the absolute magnitude of τ obtained by Said and Dickey is greater than that by the latter method. The statistic $T(\hat{\rho} - 1)$ has a similar pattern in general. One implication of the result is that the testing statistics seem to be affected by nonsensical estimates of the moving average term, and, therefore, the estimation method incorporating the Kalman filter algorithm may work better than the estimation method suggested by Said and Dickey in dealing with cases in which the parameters have a near-redundancy.

Schwert (1987, 1988) employs nonlinear least squares estimation in his simulation study, which may not be what Said and Dickey intended. The details of the computations in his studies are not reported. While investigating the size of the Said-Dickey tests, he found that, for large negative values of the moving average parameter, say -.8, the size of the ARIMA (1,0,1) test is above the nominal size based on the Dickey-Fuller distribution. He argues that the distinction between the one-step method employed by Said and Dickey and the iterative method he used seemed to account for the differences in the results over the size of the test. In his study, empirical sizes for the 5 percent level test based on Dickey-Fuller distribution of $T(\hat{\rho}_\mu - 1)$ and $\hat{\tau}_\mu$ are interpolated around .036 and .058 when the data are generated as ARIMA (0,1,1) with moving average coefficient -.8 and length 150. In a similar setting, the simulation results, obtained by incorporating the

Kalman filter for 1,000 replications, are that the empirical size of a 5 percent level test of using the statistics $T(\hat{\rho} - 1)$, $\hat{\tau}$, $T(\hat{\rho}_\mu - 1)$, and $\hat{\tau}_\mu$ are .054, .076, .118, and .069 percent respectively. These empirical sizes seem to be quite close to those obtained by Said and Dickey using true parameter values as initial values for one-step estimation with sample size 100. As is shown, the one-step method and the iterative method properly employed can produce a consistent outcome.

It is found that Said-Dickey tests also deviate from the distribution of Dickey-Fuller tests, but when the degree of discrepancy is considered, the Said-Dickey tests appear much more reliable than the Phillips-Perron tests in finite samples. But, it is notable that the results depend on the method used for the estimation of the moving average parameters; thus an accurate calculation of the related statistics sometimes turns out to be difficult.

The Lagrange Multiplier (LM) Tests by Solo

Solo (1984) extends the testing method of Dickey and Fuller (1979) in a different way based on the Lagrange Multiplier method. The LM tests require estimating only under the null hypothesis. Solo has pointed out that this feature is in marked contrast with the Said-Dickey tests where the basic theory is carried out in a setting of the Wald test.

Implementation of LM tests for a unit root is basically the same as the usual application of LM tests. First, estimated residuals under the null hypothesis, $\hat{u}_t(\Psi)$, is obtained from the restricted null hypothesis. Second, partial derivatives of all the restricted

parameters under the null hypothesis with respect to the error term u_t, $-\partial u_t/\partial\Psi$, are calculated. Third, the estimated residuals $\hat{u}_t(\Psi)$ are regressed on $-\partial u_t/\partial\Psi$ without an intercept term. Then, the LM statistics are defined as sample length times the R-square regression statistics, TR^2. The main difference of the test is that, unlike usual applications of the LM, TR^2 is not distributed as Chi-square. Solo demonstrated that TR^2 converges asymptotically to τ^2 when the mean is not considered, and to τ_μ^2 when the nonzero mean is subtracted. Accordingly, the table of the empirical distribution listed in Fuller (1976) can be used for the LM tests in a large sample. Compared with previous testing methods, the demerits of the LM approach to unit roots testing are obvious. The number of testing statistics available is decreased as the statistics of normalized first-order autocorrelation coefficients are no longer used. As squared critical values are used and it is difficult to distinguish left tail from right tail when they are squared, some difficulty lies in applying one-tailed tests using the distribution τ^2. Likewise, right tail test employing the distribution τ_μ^2 appears to be unreliable. Thus, the LM tests for a unit root become less attractive.

But an investigation is pursued for the test size of the left tail test of the LM statistic, employing the distribution of τ_μ^2. The model considered is the following,

$$y_t = \rho y_{t-1} + (1-\rho L)\mu + u_t + \theta u_{t-1},$$

where μ represents an intercept term which disappears under the null hypothesis of $\rho = 1$, but remains under the alternative. A series of

synthetic data of ARIMA (0,1,1) model with ρ = -.8 and length 150 was generated. Residual series are obtained by fitting MA (1) after the first-differencing transformation on the data, and serve as the left hand side variable. For the right hand side variable, the derivative of the error term with respect to ρ is taken, is advanced one period, and the sign is changed. Actually, obtaining the variable becomes somewhat tedious when an intercept is considered. The original series needs to be transformed by subtracting the estimate for the intercept term μ, the sample mean of the original series. With the estimate for the moving average coefficient obtained under the null, one time period forwarded values of the first derivative are produced by fitting MA (1) on the mean-subtracted data. The residual series is regressed on the one time period lagged values of the produced derivatives, and TR^2 is obtained. The empirical distribution of the LM tests, with a mean subtracted, obtained by 1000 replications appears to have test sizes much smaller than the corresponding nominal sizes. At 5 and 10 percent nominal levels, the critical values of the LM test are 8.335 and 6.641 according to the values listed in the table by Fuller (1976). The test sizes are 2.6 and 5.8 percent at 5 and 10 percent nominal levels. Thus, in the LM test with a mean subtracted, the critical values from Fuller (1976) may tend to favor the null hypothesis more than they should for a given time series.

In this chapter, it was shown that each of the Dickey-Fuller line of unit roots testing method suffers from problems of one type or another.

CHAPTER III
TESTS BASED ON INFORMATION CRITERIA

Background

In the previous chapter, features of the three types of unit roots testing methods which share all or some of the limiting distributions of the Dickey-Fuller tests have been investigated. As the unit roots testing problem in an ARIMA model is equivalent to the problem of deciding the order of integration, another three order selection criteria, whose performance can be compared to those of the preceding unit roots testing methods, are available. They are Akaike's information criteria (AIC), the Bayesian criteria of Schwarz (SBC), and the consistent order selection rule of Hannan and Quinn (HQ).

The basic idea of AIC is that the performance of a model identification procedure can be evaluated by the predictability of the estimated model about the true distribution. To understand the essential points, the derivation of AIC is sketched. Assume that one observes $x_1, \ldots, x_t$ of which the true density function is represented by $f(x|\Phi_o)$ containing the true values of the parameters. The future realization is also from this true density. Then, there exists a family of parametric density functions $f(x|\Phi)$ called predictive density functions which may differ in functional form and parameter restrictions. As a criterion of measuring how close a predictive density $f(x|\Phi)$ is to the true density $f(x|\Phi_o)$, Akaike employs a measure

of sensitivity difference:

$$I(\Phi_o;\Phi) = S(\Phi_o;\Phi_o) - S(\Phi_o;\Phi)$$

$$= \int \log f(x|\Phi_o)f(x|\Phi_o)\ dx - \int \log f(x|\Phi)f(x|\Phi_o)\ dx,$$

which is known as Kullback-Leibler mean information (KLI) or entropy. As $I(\Phi_o;\Phi) \geq I(\Phi_o;\Phi_o) = 0$, a natural target is to find $f(x|\Phi)$ that minimizes KLI. It is noted that only the mean log-likelihood $S(\Phi_o;\Phi)$ is affected by $f(x|\Phi)$. In KLI, the parameter Φ is replaced by a maximum likelihood estimator Φ_{ML}, and the quantity for deriving AIC when Φ_{ML} is sufficiently close to Φ_o becomes

$$E\ 2TI(\Phi_o;\Phi_{ML}) \simeq T(\Phi_{ML} - \Phi_o)'R(\Phi_{ML} - \Phi_o)$$

$$\simeq T(\Phi^* - \Phi_o)'R(\Phi^* - \Phi_o) + k,$$

where Φ^* belongs to the parameter subspace that does not include Φ_o, but maximizes $S(\Phi_o;\Phi)$ and R represents the Fisher information matrix. For large T, $T(\Phi_{ML} - \Phi^*)'R(\Phi_{ML} - \Phi^*)$ asymptotically converges to Chi-square distribution under some regularity conditions and thus the k represents the degrees of freedom equal to the dimension of the parameter subspace. The term $T(\Phi^* - \Phi_o)'R(\Phi^* - \Phi_o)$ is approximated, as a sample analog, by $2\{\Sigma \log f(x_i|\Phi_o) - \Sigma \log f(x_i|\Phi_{ML})\}$ plus the correction for the bias k, as Φ_{ML} is used in the place of Φ_o in the second term. Accordingly, E $2TI(\Phi_o;\Phi_{ML})$ is approximated by

$$2\{\sum_{i=1}^{T} \log f(x_i|\Phi_o) - \sum_{i=1}^{T} \log f(x_i|\Phi_{ML})\} + 2k,$$

and finally by ignoring the first term, which is common to every model, the AIC of Φ_{ML} is defined as

$$AIC(\Phi_{ML}) = (-2)\log\ (\text{maximum likelihood}) + 2k.$$

The above quantity is sufficient to measure the distance between the fitted model and true model and serves as an automatic decision rule called the minimum Akaike information theoretic decision criterion estimator (MAICE). (For details, refer to Akaike [1974, 1985].) The criterion will select the model whose calculated AIC quantity has the minimum, among those of given competing models, as the best model approximating the true model.

The AIC considers two different measures at the same time. The maximum likelihood part measures the goodness of fit that directly reflects the precision of the estimates and the bias adjustment part measures the penalty for increasing the number of parameters in a model. It is well known that there exists a trade-off between the goodness of fit and the principle of parsimony in general. Thus a selection of a model is basically the problem of choosing an appropriate combination on the trade-off. A merit of AIC is that it does not heavily rely on subjective judgement in the selection process.

There have been some controversies over the validity of AIC, which led to further development of statistical quantities for similar purposes. In the Bayesian point of view, the AIC is known to have a less firm theoretical foundation. Leamer (1979) argues that maximizing information in AIC is essentially the same as estimating with quadratic loss and the quadratic loss implies an estimation problem rather than a

model selection problem. One negative result from the argument is that the penalty for increasing the number of parameters is not unique and is illusive. (An exposition of this argument is in Amemiya [1980].) Also, as the model with a minimum risk cannot be selected when the true parameter Φ_0 is unknown and only some discrete points in the model space are considered, the resulting estimator of AIC is inadmissible.

A response from Chow (1981) points out, in favor of AIC, that while theoretically the Bayesian estimator defined by different prior densities on the true parameter Φ_0 is admissible, practically because of problems involved in the Bayesian procedure Leamer (1979) can suggest no alternative that dominates AIC for estimating Φ_0. Chow notes that the information criterion and the posterior probability criterion have distinct purposes. The former tells which model predicts the future density function better in a given sample and the latter tells which model with its prior density has the highest probability of being correct with the sample. Thus, as long as predictability of a model remains a criterion for model selection, the AIC has its own justification.

Since the emergence of AIC, many variants of information criteria have been suggested. The criterion by Sawa (1978), the Bayesian information criterion by Akaike (1978) and the Bayesian criterion of Schwarz (1978) are a few examples. In econometrics, those information criteria have been employed for resolving model selection problems. But in the time series analysis literature, more extensive studies have been made in connection with model identification problems.

In time series analysis, one of the difficult problems in identifying the data-generating process for a given time series is to decide the proper orders of an ARMA (p,q) model. The difficulty in applying the Box-Jenkins type (1976) procedure comes from the observation that the model identification stage requires experience and subjective judgement. As the information criteria provide an automatic decision rule, several quantities have been suggested for that purpose. Together with AIC, those are the SBC and HQ.

$$AIC(p,q) = (-2)\log\ (\text{max. likelihood}) + 2(p + q).$$

$$SBC(p,q) = (-2)\log\ (\text{max. likelihood}) + (p + q)\log T.$$

$$HQ(p,q) = (-2)\log\ (\text{max. likelihood}) + c(p + q)\log \log T,$$

where c is a constant greater than or equal to 2. While the SBC is derived from a Bayesian method, the result does not depend on any a priori distribution. The three quantities differ in defining a term for the penalty of increasing the number of parameters and have different statistical properties.

Shibata (1976) shows that when (p,q) are finite, estimated p and q from a minimum AIC (p,q) are not consistent and the AIC (p,q) can asymptotically overestimate the order with a positive probability. Theoretically, while the AIC does not provide consistent estimates for the order (p,q), still these estimates seem to have good one-step-ahead predictability according to Hannan (1982). Hannan (1980) shows that a minimum SBC (p,q) produces a weakly consistent estimate of the true order (p,q) as log T goes to infinity and (log T)/T converges to zero as

the observations increase infinitely. He also shows a minimum HQ (p,q) produces a strongly consistent estimator as c(log log T) decreases faster than (log T)/T. Findley (1985) mentions that, in some more realistic situations, consistency can be an undesirable property in connection with selecting model orders, and that the AIC can be more suitable. When the true model has infinitely many unknown parameters (p,q), the minimum AIC which selects models for prediction or spectrum estimation in an optimal way, has an asymptotically efficient estimator as Taniguchi (1980) and Shibata (1980, 1981) have established, but the SBC and HQ which obtain consistent estimates lead to unboundedly large losses.

Despite the arguments, the usefulness of each criterion would be mainly determined by the performances in a specific application. Thus, the application of each criterion to the selection of the order of integration, equivalently the number of unit roots, in an ARIMA model is investigated.

The Normalization of AIC by Ozaki (1977) and an Extension

Ozaki (1977) applied the AIC to the general order selection problem of an ARIMA model as a way of avoiding the difficulties involved in the identification stage of the Box-Jenkins (1976) procedure. When the order of integration is d, so that differencing the data d times is required before estimation to produce stationary series, the number of data points decreases by one every time a differencing is done. He points out that the differencing procedure changes the likelihood quantity and distorts the proper weight given to the penalty for

increasing the number of parameters and, therefore, a normalization of AIC becomes inevitable for the purpose of a proper comparison.

The Gaussian log-likelihood of an ARMA (p,q) model is expressed as

$$\log L(y;\rho,\theta,\sigma^2) = (-.5T)\log \sigma^2 - \{S(\rho,\theta)/2\sigma^2\} + h(\rho,\theta) - .5T\log 2\pi,$$

where $S(\rho,\theta) = \Sigma u_t^2$, and $h(\rho,\theta)$ is a function of ρ and θ that can be ignored in a large sample. Since an estimate of the variance is $\hat{\sigma}^2 = (1/T)\Sigma \hat{u}_t^2$, the AIC of an ARMA model is defined as

$$T\log \hat{\sigma}^2 + 2(p + q + 1) + T\log 2\pi + T.$$

When differencing d times is considered, the normalized AIC for the ARIMA (p,d,q) model may be written in the form

$$\begin{aligned} \text{AIC}(p,d,q) &= \{T/(T-d)\}\{-2\log(\text{likelihood})\} + \\ &\qquad \{T/(T-d)\}2(p + q + 1 + s) \\ &= T\log \hat{\sigma}^2 + \{T/(T-d)\}2(p + q + 1 + s) + T\log 2\pi + T, \end{aligned}$$

where s is 1 if an intercept is included, and 0 otherwise. The above definition differs slightly from Ozaki (1977) in that the use of an intercept term is not restricted, giving more flexibility. Sometimes it may need to include an intercept term even when an ARMA model is fitted on differenced series, and to omit an intercept term even when an ARMA model is estimated on undifferenced series.

Employing the same line of reasoning, the above discussion can be extended to SBC and HQ which are also used for order selection. The normalized SBC and HQ are given by

$$SBC\ (p,d,q) = \{T/(T-d)\}\{-2\log\ (\text{likelihood})\} +$$

$$(p + q + 1 + s)\log T,$$

$$= T\log \hat{\sigma}^2 + (p + q + 1 + s)\log T + T\log 2\pi + T,$$

and

$$HQ\ (p,d,q) = \{T/(T-d)\}\{-2\log\ (\text{likelihood})\} +$$

$$(p + q + 1 + s)c\log \log T,$$

$$= T\log \hat{\sigma}^2 + (p + q + 1 + s)c\log \log T + T\log 2\pi + T,$$

where s is 1 if an intercept is included, 0 otherwise and c is a constant greater than or equal to 2. A model with a minimum quantity will be selected during the order selection process.

An Application of the Normalized Information Criteria to Testing for Unit Roots

The application of the normalized information criteria to the units roots testing problem is rather straightforward. Suppose one is interested in testing whether there is a unit root in the autoregressive part of an ARMA (p,q) model. If it is known that the null hypothesis of a unit root is true, the data will be transformed by differencing once and then an ARMA (p-1,q) model will be fitted. On the contrary, if the alternative hypothesis that no unit root exists is true, then an ARMA (p,q) model would be fitted. Under the null hypothesis, the number of data points decreases by one, and the order of the autoregressive part

also decreases by one. It is noted that the three normalized information criteria have a relevance to this kind of problem.

To investigate the size of each test using these information criteria, a Monte Carlo experiment was conducted. By using a pseudo random normal variate, 1,000 replications of ARIMA (0,1,1) processes with sample length 150 and a moving average coefficient of -.8 are generated. Employing a conditional maximum likelihood estimation method, ARIMA (0,1,1) and ARIMA (1,0,1) models are estimated for each set of data, and the normalized quantities of AIC, SBC, and HQ are calculated. Those quantities are compared and the model which has the smallest magnitude are chosen. The sizes of the tests (type I errors) of the normalized AIC, SBC, and HQ are shown to be 13.2, 2.6, and 6.6 percent respectively. Comparing with the nominal size of 1 or 5 percent normally suggested with the Dickey-Fuller type tests, the normalized AIC appears to have somewhat larger size of the test. The details of the simulation process and some power studies will be discussed in the next chapter.

An Application of the Normalized Information Criteria to Distinguishing TS from DS Models

The task of distinguishing between the trend stationary (TS) and difference stationary (DS) models discussed in Nelson and Plosser (1982) in connection with the unit roots testing may be better handled using the mentioned order selection criteria. The TS and DS models are given respectively by

$$y_t = a + bt + u_t, \tag{3.1}$$

$$\nabla y_t = \mu + u_t, \tag{3.2}$$

where $u_t = \rho(L)^{-1}\theta(L)\epsilon_t$ is a stationary and invertible ARMA process. One model cannot be distinguished from the other by directly comparing the order selection quantities after estimating the models (3.1) and (3.2), as each model has a different right hand side variable. A simulation study reveals that, in general, the model with a time trend variable tends to have the smaller quantities than the other when both models fit the data seemingly equally well. During unit roots testing, equation (3.2) is likely to be estimated. When a root contained in the estimated moving average part is somewhat close to 1, one may suspect that it might be the result of differencing the model of (3.1). This type of over-differencing is believed to occur, as usually a differencing transformation of the data is recommended for removing nonstationarity in practice. Especially, it has been observed that estimating the moving average part in small samples by nonlinear optimization sometimes produces boundary estimates when the root of the true moving average coefficient is within the invertibility region. Theoretically, the probability that a local maximum can occur on the boundary point has been studied by Sargan and Bhargava (1983) for MA (1) models and Anderson and Takemura (1986) for higher order moving average models. A general consensus from many Monte Carlo studies is that, in small samples, conditional maximum likelihood estimator and conditional least squares method are preferable when the true moving average parameter is within the invertibility region. The conditional maximum likelihood (ML) method is suggested because the least squares method,

unlike ML, can lead to estimates outside the admissible parameter space. (Refer to Judge et al. [1985], pp. 302-304 for a survey of the properties of estimators.)

Nelson and Plosser (1982) try to solve the over-differencing problem by employing one of the unit root testing method of Dickey and Fuller as they find the problem of testing for a unit root in the moving average part is more discouraging because of the problems in estimation as examined in an earlier study of Plosser and Schwert (1977). The Dickey-Fuller tests are known to have low power against the alternative hypothesis of a TS model. Accordingly, it is expected that the DS model is not rejected in a vast majority of cases when the approach of Nelson and Plosser is pursued.

While a different method is illustrated for one case, the same approach can be extended to more general cases. Suppose the following ARIMA (1,1,1) model with an intercept was chosen after the selection process discussed earlier.

$$z_t = \mu + \alpha_1 z_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}, \tag{3.3}$$

where $z_t = y_t - y_{t-1}$, and the moving average parameter θ_1 is close to -1, say somewhere between -1 and -.9 and μ is a non-zero intercept. The estimated autoregressive part is assumed to have a stationary root and there is no occurrence of a common root between the autoregressive and moving average parts. The estimated moving average coefficient itself reveals nothing about which of the two, TS and DS, is appropriate. Therefore, the TS models need to be evaluated, and a model with an

appropriate ARMA error should be chosen. The order selection of ARMA error is supposed to be made by a usual application of the information criteria. Once a TS model is selected, then the model can be transformed into a first-differenced form. If the transformed model is not equivalent to (3.3) in order, the DS model is likely to explain the data better than the TS model. Otherwise, the TS and DS models are still competing. A final decision can be made by analyzing further which model has estimated errors close to the normality assumption. Diagnostic checking procedures suggested in time series analysis will serve the purpose. (As an example, refer to Brockwell and Davis [1987], pp. 296-304.) Only when a TS model is selected, equation (3.3) is likely to be the result of over-differencing.

CHAPTER IV
COMPUTATIONAL DETAILS AND A POWER COMPARISON

Computational Details

In this chapter, the details of the computational framework used throughout this study are presented. This will allow clear-cut understanding of the procedures and facilitate comparative studies.

Most of the calculations needed for this study can not be met by using specific-purpose econometric software commercially available. One of the demerits of such programs is that they come with no source code, and do not reveal the algorithms working internally. Another common problem is that the software programs are too inflexible to accommodate the changes required. It is hard to have confidence in whether the software programs are suitably written for the purposes of this study. To eliminate those problems, all the programs for the estimation and testing needed for this research have been written and extensively tested by the author. A matrix-oriented high level language, Gauss, has been used for the purpose. While the language is based on the personal-computer and cannot be run on the main-frame computer and requires, therefore, hours of computing time, it is very flexible and compact to use. Also it is known to be highly reliable in accuracy, as it uses double precision numbers as in the Fortran language.

The generation of a pseudo random normal variate consists of two steps. Initially, a pseudo uniform random variable is generated by a

multiplicative-congruential method, a most heavily-used random number generator, by the following recursion,

$$x_{t+1} = a(x_t - c)(\text{modulo } m), \quad t = 0, 1, \ldots, T,$$

where a is multiplicative constant, x_o is the seed value, c is a constant, and m is a constant. $a = 397204094$, $x_o = 1613218064$, $c = 0$, and $m = 2^{31}-1$ were used at the start. Next, those pseudo uniform numbers are called in and fed into the algorithm of Kinderman and Ramage (1976) to be transformed to pseudo normal numbers. (For discussions of various methods, refer to Kennedy and Gentle [1980], pp. 133-147 and 201-209). To see whether the generated numbers are compatible with the normality assumption, a large sample version of the W-test developed by Shapiro and Francia (1972) is applied. A set of sub-sample data, omitting the first 50 realizations, is arranged in an increasing order, $Y_{(1)} < Y_{(2)} < \ldots < Y_{(T)}$, and the data are regressed on $\Phi^{-1}([i - .5]/T)$ and an intercept, where $i = 1, 2, \ldots, T$ and the inverse of the standard normal distribution at discrete points between 0 and 1, $\Phi^{-1}(\cdot)$, can be obtained by numerical integration on a computer. As the idea comes from the relation,

$$E\, Y_{(j)} = \mu + \sigma E\, X_{(j)},$$

where $X_{(j)}$ are the sample order statistics from a standard normal distribution of size T, R^2 can be used for checking normality. By an interpolation from the values in Brockwell and Davis (1987) p. 304, it becomes that $p(R^2 < .982) \simeq .5$ and $p(R^2 < .985) \simeq .1$ for $T = 150$. Additionally, it needs to be tested whether the mean is zero using the

1.64 times the standard deviation of the sample mean, $1.64/\sqrt{T}$, since the W-test only checks normality. As the unit roots testing methods are independent of the variances of the data, the variances are not tested. The above procedure is repeated as many times as required for the creation of synthetic data sets. Once a scheme is set for an ARIMA (p,d,q) model, those realized numbers will be plugged into the model sequentially with y_{-50} and u_{-50} fixed at zero. To eliminate the start-up effect, the first 50 data points are discarded. If artificial data representing a TS model with ARMA errors are to be generated, then the deterministic part is superimposed on the ARMA processes already created.

The choice of an estimator appears to be very important in estimating an ARMA model of which an autoregressive parameter is close to unity and a moving average parameter is close to unity, as the properties of various estimators are known to differ more in those boundary cases. Plosser and Schwert (1977) and Schwert (1988) employed an iterative nonlinear least squares algorithm to estimate those ARMA models. But, in this study, an exact conditional maximum likelihood estimation method is employed, which incorporates the Kalman filter algorithm proposed in Harvey and Phillips (1979) and Harvey (1981). The rationale is that the Kalman filter provides an optimal solution to the problem of prediction and updating, and the likelihood function of an ARMA model can be decomposed in terms of prediction errors. A merit of the algorithm is that the inversion of the variance-covariance matrix can be completely avoided. First a general ARMA (p,q) model is written

in the form of a state space model. A state space model consists of measurement equation (4.1) and transition equation (4.2). The measurement equation is given by

$$y_t = z_t'\alpha_t + \xi_t, \qquad t = 1, 2, \ldots, T, \tag{4.1}$$

where y_t is observed, state vector α_t cannot be observed directly, z_t is a row vector consisting of one and zeros and ξ_t is normally distributed with zero mean and variance $\sigma^2 h_t$. Then, the state vector α_t is generated by the process

$$\alpha_t = T\alpha_{t-1} + R\eta_t, \qquad t = 1, 2, \ldots, T, \tag{4.2}$$

where T is a transition matrix, R is a matrix of fixed coefficients and η_t is distributed as normal with zero mean and variance $\sigma^2 Q$. Prediction step uses the equations (4.3) and (4.4),

$$a_{t|t-1} = Ta_{t-1}, \qquad t = 1, 2, \ldots, T, \tag{4.3}$$

where a_{t-1} represents the minimum mean squares estimator of α_{t-1} given all the information up to t-1 period, and

$$P_{t|t-1} = TP_{t-1}T' + RQR', \qquad t = 1, 2, \ldots, T, \tag{4.4}$$

where $\sigma^2 P_{t-1}$ denotes the covariance matrix of $a_{t|t-1}$. Then the prediction error (4.5) of the observed variable and the error variance $\sigma^2 f_t$ (4.6) are obtained as follows:

$$v_t = y_t - z_t'a_{t|t-1}, \qquad t = 1, 2, \ldots, T, \tag{4.5}$$

$$f_t = z_t'P_{t|t-1}z_t + h_t, \qquad t = 1, 2, \ldots, T. \tag{4.6}$$

With the arrival of a new observation, the estimator for α and P matrix is updated as in (4.7) and (4.8),

$$P_t = P_{t|t-1} - P_{t|t-1}z_t z_t' P_{t|t-1}' / f_t, \quad t = 1, 2, \ldots, T, \tag{4.7}$$

$$a_t = a_{t|t-1} + P_{t|t-1}z_t(y_t - z_t' a_{t-1})/f_t, \quad t = 1, 2, \ldots, T. \tag{4.8}$$

After processing T observations, T prediction errors, each of which is independently and normally distributed with zero mean and variance $\sigma^2 f_t$, will be available. (For a complete derivation of the above equations, refer to Harvey [1981], Chapter 4.) Accordingly, the exact log-likelihood is maximized over the autoregressive parameter ρ and the moving average parameter θ as in

$$\ln L(\rho,\theta) = -.5T[\ln(2\pi) + 1] - (.5T)\ln \hat{\sigma}^2 - (.5)\ln \Sigma f_t, \tag{4.9}$$

where $\hat{\sigma}^2 = T^{-1} \Sigma v_t^2/f_t$. As h_t is suppressed to zero in an ARMA setting, only elements of a_o, P_o, and starting parameter values are required. In practice, zero vector works fine as a_o, and P_o can be easily obtained by the relation (4.10) as long as the roots of the autoregressive part are less than 1. But, when P_o needs to be evaluated for a nonstationary ARMA model, one can no longer rely on the relation

$$\text{vec}(P_o) = [I - T\otimes T]^{-1}\text{vec}(RQR'), \tag{4.10}$$

because it breaks down when the system is nonstationary. In this study, the P_o matrix of the ARMA models, where at least one of the autoregressive roots is right on or slightly inside the unit circle, needs to be supplied. Even when a true model is a stationary one, the

situation can happen during a nonlinear optimization process that at least one of the roots crosses the unit circle border from outside by the force of the step length and direction. It is uncertain whether there exists any method that will work completely satisfactorily for the situation. Though it looks like an art rather than a science, Harvey (1981) suggests that the recursion begin at $t = 0$, with a zero vector for a_o and $P_o = wI$, where w is a large number. But in forming the likelihood function, only prediction errors for $t > m$ is considered. In practice, it may be difficult to choose m and w in an optimal way. Some experiments performed by the author did not apparently justify the method, as the convergence often failed, or was very slow compared to the method employing entire prediction errors for $t > 0$. Therefore, a maximum likelihood estimation program is written in such a way that whenever the situation occurs, $P_o = (10^{16})I$ is simply substituted in place of P_o calculated from the relation (4.10). The likelihood and the parameter estimates do not seem to be noticeably affected by the method. Another finding is that the summation of the natural log values of f_t, the quantity related to prediction error $\sigma^2 f_t$, is significantly large in the ARMA (1,1) models considered. Thus, minimizing over only the conditional sum of squares may produce a less accurate result. Especially, when the moving average coefficient θ of ARMA (1,1) or MA (1) model bounces out of an invertible range, the objective function is switched from (4.9) to the following by taking a reciprocal of θ,

$$\ln L(\rho, 1/\theta) = -.5T[\ln(2\pi) + 1] - (.5T)\ln(\theta^2\hat{\sigma}^2) - (.5)\ln \Sigma f_t, \qquad (4.10)$$

where $\hat{\sigma}^2 = T^{-1} \Sigma v_t^2/f_t$. The switching does no harm and, in some cases, appears to help in searching for an optimal point more efficiently.

The likelihood function should be optimized by employing a nonlinear optimization method. A general consensus is that an appropriate combination of a nonlinear optimization method and a step length search algorithm is demanded for a specific type of objective function. A quasi-Newton nonlinear optimization method is employed, which is based on the symmetric positive definite secant update algorithm, what is known as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update, combined with the golden section search algorithm for step length. (For a detailed explanation of BFGS update, refer to Dennis and Schnabel [1983], Chapter 9, and for the golden section search algorithm, see Kennedy and Gentle [1980], pp. 432-433). Actually, the Hessian is computed only twice, at the beginning and after convergence to obtain an accurate standard deviation. Gradients and Hessians are numerically evaluated.

Supplying a starting value close to the true coefficients is very important for the success of optimization and different starting values should be tested as a check against a local convergence. When a unit root is suspected in the autoregerssive part of an ARMA model, it would be better to transform the data by first-differencing and apply a method such as an innovation algorithm to find preliminary estimates only for the stationary part. Especially when an unrestricted ARMA model is estimated with the original data, then a rewarding strategy is to supply a preliminary autoregressive coefficient, of which the root is slightly

outside a unit circle, say .995, together with the preliminary moving average estimates, obtained previously from the transformed data, as a starting parameter vector. Care should be taken when there is a possibility of a near redundancy in parameters.

A Simulation Study of the Power of the Tests

In the previous two chapters, the Dickey-Fuller line of unit roots tests and information criteria for the order selection have been addressed. As they have different theoretical backgrounds, it may be hard to assess those tests together in a general environment. But, at least, in testing for the presence of unit roots in an ARMA model where errors are behaving nicely, the relative power of each test may be compared without much difficulty. Here, an investigation is made to answer the question of whether the Dickey-Fuller line of tests is any better than the order selection tests.

The empirical distributions and their critical values tabulated in Fuller (1976) are frequently used as a yardstick, apparently without hesitation. Whatever the reasons are, it is still a question whether those critical values are reliable when a different simulation process is taken. To investigate the problem, 4,000 random walk series with length 150 have been generated following the method previously elaborated. Note that the initial 50 observations have been discarded to eliminate the start-up effect in the simulation. Table 4-1 reports the normalized first-order autocorrelation coefficient and the t-test with and without an intercept respectively, and the Dickey-Fuller critical values linearly interpolated. The statistics $T(\hat{\rho}_\mu - 1)$, $\hat{\tau}$, and $\hat{\tau}_\mu$ appear to be, relatively speaking, close to the critical values

Table 4-1
The comparison of critical values listed in Fuller (1976) and the ones obtained by a different simulation.

nominal size	D-F value	actual size	D-F value	actual size
	$T(\hat{\rho} - 1)$		$\hat{\tau}$	
2.5 %	-10.23	1.50 %	-2.237	2.73 %
5.0 %	- 7.93	3.30 %	-1.95	5.35 %
10.0 %	- 5.63	7.35 %	-1.613	10.50 %
	$T(\hat{\rho}_\mu - 1)$		$\hat{\tau}_\mu$	
2.5 %	-16.4	2.40 %	-3.16	2.53 %
5.0 %	-13.8	5.25 %	-2.887	5.20 %
10.0 %	-11.07	10.90 %	-2.577	11.40 %

Note: The tabulation reflects the result of 4,000 replications of data with length 150. The critical values of Dickey-Fuller table were linearly interpolated.

of Dickey and Fuller. On the other hand, the statistics $T(\hat{\rho} - 1)$ shows notable differences at all the three nominal significance levels. The actual size of the test turned out to be systematically lower than is claimed. One implication of the result is that $T(\hat{\rho} - 1)$ is relatively more sensitive to the method of generating synthetic data than the other, and, therefore, it may be inadequate to use without proper knowledge about the data-generating process.

It is noted that the estimated sizes of each test investigated in Chapter II by employing some nominal levels of the Dickey-Fuller critical values show noticeable discrepancies as listed in table 4.2 of the simulation results. An ARIMA (0,1,1) with a large moving average coefficient, -.8, was generated and the estimated test sizes were

obtained respectively at the nominal 2.5, 5, and 10 percent levels. The Phillips-Perron tests tend to reject too frequently the null hypothesis of the presence of a unit root, when the truncation point L is fixed at either 4 or 12, as pointed out for other Dickey-Fuller line of tests in Schwert (1988). The estimated size of the Said and Dickey tests, after 25 iterations, is generally larger than the nominal sizes. The LM test appears to have smaller estimated sizes than the nominal sizes. Accordingly, the empirical distributions of Dickey and Fuller for a critical value cannot be relied on. The information criteria being considered in this study do not use any fixed level critical value for the testing. In order to establish a standard for a comparison of the power, therefore, the estimated test sizes of the three information criteria under the null hypothesis are obtained first and then comparable critical values of each of the Dickey-Fuller line of tests at those sizes are newly constructed from the empirical distributions constructed by simulations. Table 4-3 reports the comparative power of each unit roots test at three nominal levels when the true model is an ARIMA (1,0,1) with $(\rho,\theta) = (.95,-.8)$ and sample length 150. The critical value employed at each nominal level was empirically obtained from an ARIMA (0,1,1) model with $\theta = -.8$ and sample length 150. Obviously, when the size of the type I error is fixed at 13.2, 2.6 and 6.6 percent consistently for each test, two of the Phillips-Perron tests, $\hat{\tau}$ and $T(\hat{\rho} - 1)$ with a fixed $L = 4$, outperform the information criteria by about a 3 to 7 percent margin. But the information criteria outperform those two Phillips-Perron tests by about a 1 to 14 percent margin when the truncation number L is fixed at 12 instead of 4.

Table 4-2
The estimated test sizes of the Dickey-Fuller line of unit roots tests when the critical values of the Dickey and Fuller table are used when the model is an ARIMA (0,1,1) with $\theta = -.8$ and sample length 150.

method of	test used	nominal size	obtained value	estimated size
Phillips	$T(\hat{\rho} - 1)$	2.5 %	-10.23	73.9 %
and	with	5.0 %	-7.93	79.2 %
Perron	$L = 4$	10.0 %	-5.63	84.2 %
	$T(\hat{\rho} - 1)$	2.5 %	-10.23	82.3 %
	with	5.0 %	-7.93	85.9 %
	$L = 12$	10.0 %	-5.63	89.3 %
	$\hat{\tau}$	2.5 %	-2.24	75.8 %
	with	5.0 %	-1.95	80.6 %
	$L = 4$	10.0 %	-1.61	85.4 %
	$\hat{\tau}$	2.5 %	-2.24	82.9 %
	with	5.0 %	-1.95	86.2 %
	$L = 12$	10.0 %	-1.61	90.1 %
Said		2.5 %	-10.23	3.0 %
and	$T(\hat{\rho} - 1)$	5.0 %	-7.93	5.4 %
Dickey		10.0 %	-5.63	9.7 %
		2.5 %	-2.24	4.0 %
	$\hat{\tau}$	5.0 %	-1.95	7.6 %
		10.0 %	-1.61	13.3 %
LM	TR^2 obtained	2.5 %	9.99	1.3 %
by	with	5.0 %	8.33	2.6 %
Solo	a constant	10.0 %	6.64	5.8 %

Note: Estimated sizes were obtained from 4,000 replications for the Phillips-Perron tests and 1,000 replications for others. The Dickey-Fuller critical values at each nominal level were linearly interpolated from the table in Fuller (1976).

Table 4-3
The comparative power of each unit roots test at three nominal levels when the true model is an ARIMA (1,0,1) with $(\rho,\theta) = (.95,-.8)$ and sample length 150.

method of	test used	nominal size	obtained value	estimated size
Phillips	$T(\hat{\rho} - 1)$	2.6 %	-129.27	40.1 %
and	with	6.6 %	-108.99	75.8 %
Perron	L = 4	13.2 %	-90.53	93.2 %
	$T(\hat{\rho} - 1)$	2.6 %	-201.88	22.8 %
	with	6.6 %	-179.89	49.0 %
	L = 12	13.2 %	-156.27	78.2 %
	$\hat{\tau}$	2.6 %	-9.72	44.0 %
	with	6.6 %	-7.93	78.5 %
	L = 4	13.2 %	-7.69	94.0 %
	$\hat{\tau}$	2.6 %	-11.26	31.5 %
	with	6.6 %	-10.41	65.7 %
	L = 12	13.2 %	-9.51	89.7 %
normalized	SBC	2.6 %	n.a.	36.8 %
information	HQ	6.6 %	n.a.	71.2 %
criteria	AIC	13.2 %	n.a.	90.6 %
Said		2.6 %	-11.27	28.8 %
and	$T(\hat{\rho} - 1)$	6.6 %	-7.06	47.8 %
Dickey		13.2 %	-4.44	63.4 %
		2.6 %	-2.44	21.6 %
	$\hat{\tau}$	6.6 %	-2.03	40.4 %
		13.2 %	-1.62	58.8 %
LM	TR^2 obtained	2.6 %	8.52	5.8 %
by	with	6.6 %	8.41	24.3 %
Solo	a constant	13.2 %	6.77	46.8 %

<u>Note</u>: Estimated sizes were obtained from 4,000 replications for the Phillips-Perron tests and 1,000 replications for others. The nominal test sizes were set according to the size of type I error of normalized SBC, HQ, and AIC for an ARIMA (0,1,1) with $\theta = -.8$ and sample length 150. The above critical values were from empirical distributions.

Presumably, it seems that, when the Phillips-Perron tests are employed, choosing an appropriate truncation point L is important for securing the power. Seemingly, the severe finite sample deviation from the limiting distribution does not seem to greatly affect the power. If the tests are ranked according to only a power consideration, it may be that those Phillips-Perron tests should be first, the information criteria second and the Said-Dickey third. The LM test seems to be the least powerful test because it considers an intercept. It has not been tested whether the same result will follow when a different maximum likelihood estimation method is utilized. Recall that the higher power the Phillips-Perron tests exhibit is the result obtained with the complete knowledge that the moving average coefficients are the same for the two models being investigated. A discouraging factor is that, as reported in Schwert (1987), the empirical distribution of the Phillips-Perron tests changes considerably for a different moving average parameter value and the empirical values need to be tabulated for every finite sample length and model specification. As the model specification is unknown in practice, the difficulty involved in applying a correct critical value may more than offset the merit of higher power. In contrast, the information criteria do not suffer from those problems. Accordingly, the information criteria are more useful than the other tests in the aspect of power, reliability, and ease of implementation. In unit roots testing, customarily, the type I error used to be fixed at the 1 or 5 percent level. But, this study reveals that this tradition may not be justified and recommended. The reason is that the type II error appears to be very large, around 30 to 50 percent at the 5 percent

nominal level and more than 60 percent at the 1 percent nominal level, and, thus, the tests favor the null hypothesis too often when the true process is stationary. If it is not an occasion where one needs to guard more against committing type I error than type II error, it may be that the normalized AIC and HQ are more recommendable in detecting the subtle difference when an ARIMA (0,1,1) and an ARIMA (1,0,1) are competing. Otherwise, the normalized SBC is suggested.

CHAPTER V
SUMMARY AND CONCLUSIONS

The nonstationary behavior of macroeconomic series has drawn the attention of many economists in recent years for both theoretical and empirical reasons. In this circumstance, formal unit roots testing methods have been proposed in many different forms. Among them, the Phillips-Perron testing method seemed to dominate the other methods in theoretical elegance, ease of use, and very general applicability. Lately, Schwert (1987, 1988) discovered that many macroeconomic time series have a large moving average coefficient and, in the circumstances, most of the unit roots testing methods do not work properly. Motivated by his findings, an attempt has been made to find a better method that will be more reliable in practical situations. For this purpose, it has been focused on the Dickey-Fuller line of unit roots testing method and the normalized information criteria.

All of the Dickey-Fuller line of testing statistics investigated also diverge from the limiting distribution of the corresponding Dickey-Fuller testing statistics as Schwert (1987, 1988) reports. But, the deviations are less severe than the findings of Schwert as, in this study, the models with no intercept were tested and the statistics derived considering an intercept were avoided if possible.

While the Phillips-Perron tests have been claimed to have the most general applicability and have been applied as such, some unfavorable

evidence against the tests is presented. According to the investigation, in an ARMA (1,1) model with a large moving average coefficient, the calculated quantities of the testing statistics vary in a wide range of values according to the choice of the truncation point L and the lag windows. Also the autocorrelation function constructed from the least squares residuals shows no clear pattern for an optimal choice of L. By a simulation, it became clear that an arbitrary fixing of L can produce a significant difference in the power of the test. The Phillips-Perron tests do not seem to work properly when the degree of heteroscedasticy is high in a time series even in the random walk process without a drift. Employing the critical values from the tables of Fuller (1976) can be misleading in this case just as in ARMA models. Judging from the results, together with earlier findings of Schwert, it would be better to be very cautious about the interpretation of the testing results of the Phillips-Perron tests.

The Said-Dickey testing method has a disadvantage in estimation method. As there is no mechanism for securing an invertible moving average estimate, the method sometimes produced noninvertible estimates, which also adversely affected the calculation of the statistics. By using an objective function written in terms of the Kalman filter, the problem could be somewhat alleviated. The empirical size of the test derived from the estimates obtained after 25 iterations are roughly consistent with the result of Said and Dickey when they started with the true moving average value.

The LM tests of Solo turned out to be the least attractive. Because of the squared feature, one test is available for the left tail.

The empirical size of the test appeared to be about half of the nominal size.

In this study, the applicability of the information criteria has been also investigated. Despite the controversies about them, each of the three criteria, AIC, SBC, and HQ, seems to have its own justification and basis. Following the extension of AIC by Ozaki, the SBC and HQ were also normalized for identifying a unit root in an ARMA series. An application of the information criteria to distinguishing between the TS and DS models has been discussed. An exact maximum likelihood method was pursued by incorporating the Kalman filter such that a switching of the objective function for a noninvertible moving average parameter and a switching of the initial matrix P_0 for a nonstationary autoregressive parameter are made during the nonlinear estimation. The normalized AIC seems to have around 13 percent of nominal size, which is contrasted with the usual fixing of the nominal size at 1 or 5 percent.

The empirical distribution of the Dickey and Fuller tests were reconstructed with 4,000 replications of the AR (1) process. Curiously enough, the empirical distribution of $T(\hat{\rho} - 1)$ appeared to be sensitive to the process of generating artificial data. As reported, the distributions of other statistics appeared quite reliable compared with the Dickey-Fuller critical values. To compare the power, first, the critical values of the tests have been derived by simulation for the three nominal levels, which represents the sizes of the test of the normalized information criteria. When L was fixed at 4, the Phillips-Perron tests appear to have the highest power. The power of the

normalized information criteria appeared to be lower than that, but, in turn, higher than the power of the Phillips-Perron tests obtained by fixing L at 12. The Said-Dickey tests fell to third place and the LM test to last place.

If a unit roots testing method for an ARMA (1,1) model with a large moving average coefficient has to be chosen by considering the intrinsic problem and power of each test, the information criteria will be selected as the more practical and viable methods. Arranged by the order of importance of guarding against committing the type I error, the normalized SBC comes before the normalized HQ and lastly the normalized AIC. But, still, a testing result by any of the normalized information criteria should be considered as statistically indicative rather than definitive.

In this study, the number of replications may appear somewhat small. Therefore, some experiments may have room for improvement in accuracy, but no drastic change is likely even if we increase the number of replications, as the Monte Carlo simulation is very slow in the rate of convergence, in general.

BIBLIOGRAPHY

Akaike, H. (1974), "A New Look at the Statistical Identification Model," IEEE Transactions on Automatic Control, 19, 716-723.

Akaike, H. (1978), "Time Series Analysis and Control through Parametric Models," in Applied Time Series Analysis edited by D. F. Findley, Academic Press, New York.

Akaike, H. (1985), "Prediction and Entropy," in A Celebration of Statistics: the ISI Centenary Volume, edited by A. C. Atkinson and S. E. Fienberg, Springer-Verlag, New York.

Amemiya, T. (1980), "Selection of Regressors," International Economic Review, 21, 331-354.

Anderson, T. W. and Takemura, A. (1986), "Why Do Noninvertible Estimated Moving Averages Occur?" Journal of Time Series, 7, 235-254.

Baillie, R. T. and Bollerslev, Tim (1987), "Multivariate GARCH Processes and Models of Time Varying Risk Premia in Foreign Exchange Markets," Mimeo, Michigan State University and Northwestern University.

Box, G. E. P. and Jenkins, G. M. (1976), Time Series Analysis: Forecasting and Control, Revised Edition, Holden-Day, San Francisco.

Brockwell, P. J. and Davis, R. A. (1987), Time Series: Theory and Methods, Springer-Verlag, New York.

Campbell, Y. J. and Mankiw, N. G. (1987a), "Are Output Fluctuation Transitory?," Quarterly Journal of Economics, 102, 857-880.

Campbell, Y. J. and Mankiw, N. G. (1987b), "Permanent and Transitory Components in Macroeconomic Fluctuations," American Economic Review Proceedings, 77, 111-117.

Chatfield C. (1984), The Analysis of Time Series: An Introduction, 3rd Edition, Chapman and Hall, New York.

Chow, G. C. (1981), "A Comparison of the Information and Posterior Probability Criteria for Model Selection," Journal of Econometrics, 16, 21-33.

Cooper, J. P. and Nelson, C. R. (1975), "The Ex-Ante Prediction Performance of the St. Louis and F.R.B-M.I.T.-Penn. Econometric Models and Some Results on Composite Predictors," Journal of Money, Credit and Banking, 7, 1-32.

Corbae, D. and Ouliaris, S. (1986), "Robust Tests for Unit Roots in the Foreign Exchange Market," Economics Letters, 22, 375-380.

Dennis, J. E. and Schnabel, R. B. (1983), Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, New Jersey.

Dickey, D. A. and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," Journal of the American Statistical Association, 74, 427-431.

Dickey, D. A. and Fuller, W. A. (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," Econometrica, 49, 1057-1072.

Engle, R. F. and Granger, C. W. J. (1987), "Co-integration and Error Correction: Representation, Estimation, and Testing," Econometrica, 55, 251-276.

Evans, G. B. A. and Savin, N. E. (1981), "Testing for Unit Roots: 1," Econometrica, 49, 753-779.

Evans, G. B. A. and Savin, N. E. (1984), "Testing for Unit Roots: 2," Econometrica, 52, 1241-1269.

Findley, D. F. (1985), "On the Unbiased Property of AIC for Exact or Approximate Linear Stochastic Time Series Models," Journal of Time Series Analysis, 6, 229-252.

French, K., G. Schwert and Stambaugh, R. (1987), "Expected Stock Returns and Volatility," Journal of Financial Economics, 19, 3-30.

Fuller, W. A. (1976), Introduction to Statistical Time Series, John Wiley & Sons, New York.

Granger, C. W. J. (1981), "Some Properties of Time Series Data and Their Use in Econometric Model Specification," Journal of Econometrics, 121-130.

Granger, C. W. J. and Newbold, P. (1974), "Spurious Regressions in Econometrics," Journal of Econometrics, 2, 111-120.

Granger, C. W. J. and Weiss, A. A. (1983), "Time Series Analysis of Error Correcting Models," in Studies in Econometrics, Time Series, and Multivariate Statistics, Academic Press, New York.

Hannan, E. J. (1982), "Testing for Autocorrelation and Akaike's Criterion," in Essays in Statistical Science, Applied Probability Trust, New York.

Hannan, E. J. and Quinn, B. G. (1979), "The Determination of the Order of an Autoregression," Journal of the Royal Statistical Society, 41, 190-195.

Harvey, A. C. (1981), Time Series Models, John Wiley and Sons, New York.

Harvey, A. C. and Phillips, G. D. A. (1979), "Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances," Biometrika, 66, 49-58.

Hendry, D. F. (1986), "Econometric modelling with Cointegrated Variables: An Overview," Oxford Bulletin of Economics and Statistics, 48, 201-212.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., and Lee, T. C. (1985), The Theory and Practice of Econometrics, John Wiley and Sons, New York.

Kennedy, W. J. and Gentle, J. E. (1980), Statistical Computing, Marcel Dekker, New York.

Kinderman, A. J. and Ramage, J. G. (1976), "Computer Generation of Normal Random Numbers," Journal of the American Statistical Association, 71, 893-896.

Leamer, E. (1979), "Information Criteria for the Choice of Regression Models," Econometrica, 47, 507-510.

Litterman, R. B. (1986), "Forecasting With Bayesian Vector Autoregressions: Five Years of Experience," Journal of Business and Economic Statistics, 4, 25-38.

Lo, A. W. and MacKinlay, A. C. (1989), "The Size and Power of the Variance Ratio Test in Finite Samples: A Monte Carlo Investigation," Journal of Econometrics, 203-238.

Meese, R. A. and Singleton, K. J. (1982), "On Unit Roots and the Empirical Modeling of Exchange Rates," Journal of Finance, 37, 1029-1037.

Nelson, C. R. and Kang, Heejoon (1981), "Spurious Periodicity in Inappropriately Detrended Time Series," Econometrica, 49, 741-751.

Nelson, C. R. and Plosser, C. I. (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implication," Journal of Monetary Economics, 10, 139-162.

Nelson, C. R. and Schwert, G. W. (1977), "On Testing the Hypothesis that the Real Rate of Interest is Constant," American Economic Review, 67, 478-486.

Ozaki, T. (1977), "On the Order Determination of ARIMA Models," Applied Statistics, 26, 290-301.

Perron, Pierre (1986), "Trend and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach," Mimeo, Université de Montréal, Montreal.

Perron, Pierre and Phillips, P. C. B. (1987), "Does GNP Have a Unit Root?: A Re-evaluation," Economics Letters, 23, 139-145.

Phillips. P. C. B. (1986), "Understanding Spurious Regressions in Econometrics," Journal of Econometrics, 33, 311-340.

Phillips. P. C. B. and Perron, Pierre (1986), "Testing for a Unit Root in Time Series Regression," Cowles Foundation Discussion Paper No. 795.

Plosser C. I. and Schwert G. W. (1977), "Estimation of a Non-Invertible Moving Average Process: The Case of Over-Differencing," Journal of Econometrics, 6, 199-224.

Porteba, J. and Summers, L. (1986), "The Persistence of Volatility and Stock Market Fluctuations," American Economic Review, 76, 1142-1151.

Said, S. E. and Dickey, D. A. (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," Biometrika, 71, 599-608.

Said, S. E. and Dickey, D. A. (1985), "Hypothesis Testing in ARIMA(p,1,q) Models," Journal of the American Statistical Association, 80, 369-374.

Sargan, J. D. and Bhargava, A. (1983), "Maximum Likelihood Estimation of Regression Models with First-Order Moving Average Errors When the Root Lies on the Unit Circle," Econometrica, 51, 799-820.

Sawa, T. (1978), "Information Criteria for Discriminating Among Alternative Regression Models," Econometrica, 46, 1273-1291.

Schwarz, G. (1978), "Estimating the Dimension of a Model," The Annals of Statistics, 8, 1071-1081.

Schwert, G. W. (1987), "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data," Journal of Monetary Economics, 20, 73-103.

Schwert, G. W. (1988), "Tests for Unit Roots: A Monte Carlo Investigation," University of Rochester Working Paper No. GPB 87-01, Revised Version.

Shapiro, S. S. and Francia, R. S. (1972), "An Approximate Analysis of Variance Test for Normality," Journal of the American Statistical Association, 67, 215-216.

Shibata, R. (1980), "Asymptotic Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process," The Annals of Statistics, 8, 147-164.

Shibata, R. (1981), "An Optimal Autoregressive Spectral Estimate," The Annals of Statistics, 9, 300-306.

Solo, Victor (1984), "The Order of Differencing in ARIMA Models," Journal of the American Statistical Association, 79, 916-921.

Stock, J. H. and Watson, M. W. (1986), "Does GNP Have a Unit Root?," Economics Letters, 22, 147-151.

Taniguchi, M. (1980), "On the Selection of the Order of the Spectral Density Model for a Stationary Process," Annals of the Institute of Statistical Mathematics, 32A, 401-409.

BIOGRAPHICAL SKETCH

Yang-Seob Lee was born in Seoul, Korea, on March 11, 1955, two years after the Korean War ended. Life was never easy for him during the younger days. Motivated by a scholarly curiosity and social need, he determined to study economics at the Kon-Kuk University in Seoul in 1975. In the middle of the school year, he had to join the army and spent twenty-nine months in the military. After being discharged, he continued to develop an interest in the sociological, political, economic, and ecological aspects of human activities. In 1982, he received a B.A. degree by working on a study of the economic solutions for pollution problems. To broaden his knowledge and experience, he came to the United States as an international student and entered the University of North Texas. He studied development theory for the third world countries and macroeconomic theory. After obtaining an M.S. degree in economics in 1984, he transferred to the University of Florida. Ever since, he has been concentrating in the field of time series analysis, econometrics, and open-economy macroeconomics.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

G.S. Maddala, Chairman
Graduate Research Professor of Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

David A. Denslow

David A. Denslow
Professor of Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mark Rush
Associate Professor of Economics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John S. Shonkwiler
Professor of Food and Resource Economics

This dissertation was submitted to the Graduate Faculty of the Department of Economics in the College of Business Administration and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1989

Dean, Graduate School